SEQUENTIAL PROCEDURES IN MEDICINE

by

JERRY WAYNE HIGHFILL 265

B. A., Southwestern College, 1967

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics
and Computer Science

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1968

Approved by:

Major Professor

TABLE OF CONTENTS

0. Introduction

Wetherill (17) defined a sequential experiment as one in which the course of the experiment depends in some way upon the results obtained at each stage.

In a medical setting patients enter for treatment in succession. Decisions must be made for therapy. Novel treatments must be considered or medicine would stagnate. At the same time patients must be protected from receiving a dangerous medication or prophylactic. Ethical standards require a physician to use treatments which he deems best. If he is unable to distinguish between responses, he is to use what he considers the best as soon as it becomes apparent. In many cases life is at stake in the evaluation of treatments.

Experimentation with animals in laboratories has already indicated that certain treatments are effective. The researcher can only surmise from his experience how efficient a treatment will be when applied to humans. Some criterion has to be used to decide upon the effectiveness of the treatment and at the same time protect as many people as possible.

Under these ethical considerations Armitage urges the use of sequential experiments. This type of design aids the physician in his continuous scrutiny of his patients and allows decisions to be made rapidly when response differences are large.

The sequential procedure was first introduced by Wald at the end of World War II. The war had caused his work to be

considered restricted and was not released until later. Wald's work proved to be particularly useful in sampling inspection.

Armitage recognized that Wald's ideas met some of the demands of medicine. Not only did he recommend some of Wald's designs, but he helped develop some plans that he hoped would be even more suitable.

In 1957 Armitage (2) presented a restricted procedure for testing the mean of a normal population with known variance. This restricted plan was constructed by the diffusion process. Using a Monte Carlo method the author found the average sample number (A.S.N.) of the restricted procedure to be much larger than was expected. Schneiderman and Armitage (12) then introduced a wedge design. This design gave a family of plans that contained in one extreme the restricted procedure and in another extreme (when the restricted sample size $N \to \infty$) approached the open sequential procedure of known variance. The wedge design proved to have a very good A.S.N.

In 1962 Schneiderman and Armitage (13) proposed a wedge design for a closed sequential t-test. This procedure was packed with assumptions. Unlike the wedge design with known variance, whose outer boundaries coincided with the corresponding open design, the new plan only approximated the upper boundary presented for hypotheses (6.1). The authors tried very hard to make their design look like it followed as naturally as the ones where the variance was assumed to be

known, but the closed sequential t procedure with all of its approximations appears to this writer to be too much an amalgamation of (12) to be included in section 6.

This report exposes the basic considerations for designing a sequential medical trial (section 1) and presents some of the sequential methods to evaluate data, which is either qualitative (section 3) or quantitative (section 6). Wald's theory of the Sequential Probability Ratio test is presented in section 2. Section 4 deals with truncated and restricted designs of the binomial case, and section 5 discloses some applications from medicine.

1. Some Problems in the Design of Sequential Medical Trials

Armitage (3) indicated three reasons for testing sequentially: (1) economy, (2) estimation with desired accuracy (see J. Neyman's Lectures and Conferences on Mathematical Statistics and Probability), and (3) ethical considerations. Since the fixed sample test procedures with a prescribed size and power generally require large samples, any reduction in the average or expected length of a test, without sacrificing precision is usually welcome. Sequential tests on the average require a substantially smaller number of observations than do test procedures based on a predetermined number of observations (15).

The precision of an estimate depends on the sample size and the variance of the observation. Armitage (1) concluded that non-sequential estimators could be used in the binomial case even though the sampling was done sequentially.

Some of the difficulties with planning a sequential design are: (1) responses (data) may not be available as soon as desired (or within a period of time which is short in comparison with the period during which the subjects enter the trial); (2) finding the important qualities of the treatment upon which the stopping rule will depend; and (3) dealing with the complex problems which arise if the design is to include cooperative trials of many physicians or clinics.

In deciding the criterion for the stopping rule one must realize that if a decision is made to stop suddenly when a large difference is recorded that the experimenter must be prepared to lose degrees of precision in the comparisons of other responses. This idea is of overriding importance to cases where the response is based on mortality.

The uncertainty about the ultimate length of the trial may cause administrative difficulty if a special staff has to be organized or if an estimate of cost is required before the trial begins. Another difficulty is that results have to be sent to the core organization rather than just being compiled as the experiment is finished.

Most of the experimentation in medicine has been sustained by fairly simple experiments. The main reason is that in a clinic or hospital ward it is difficult enough to run the simplest designs without compounding this trouble with more controls and confusion to the participating staff and patients.

Usually sequential analyses are performed on pairs of observations where each subject has only one of the alternative treatments, and treatment comparisons are made between subjects; or each subject has more than one treatment (within subject) where the researcher assumes that the periods of treatment influence do not overlap. Comparisons between patients are usually used for acute diseases, where the treatment period is limited or the treatment takes a long time,

and for prophylactic trials; whereas, the within subject design is used when the experimenter feels that the response times are short and when it is desired to help eliminate variations from subject to subject.

Care must be taken to prevent sampling bias. It was suggested that for within patient comparisons, a natural method was to pair two successive observations on the same subject, randomizing the treatment order. For between subject comparisons the treatments can still be randomized only this time on successive subjects entering as pairs into the trial.

Armitage (3) felt that pairing did not appreciably reduce the efficiency of the design. Some efficiency may be gained if successive subjects entering the trial are more alike than subjects chosen randomly from the whole series of subjects.

If natural stratification is present, it was suggested that this separation into groups was acceptable if it would reduce the total variation, but the researcher is warned that too many sub-divisions could cause many unpaired patients and could sharply reduce the efficiency of the test.

2. Wald's Sequential Probability Ratio Test of Statistical Hypotheses

To discuss the designs which are to be introduced in the next sections, some basic concepts need to be introduced.

The procedure for testing hypotheses can be considered in the framework of the general decision problem. There are two possible terminal actions, a_1 and a_2. The appropriate action to be taken depends on an unknown parameter θ which belongs to Ω, the parameter space. Each parameter represents a distribution function of the observed random variable. The set Ω can be decomposed into sets w_1 and w_2 such that action a_1 is preferred if θ is in w_1 and a_2 is preferred if θ is in w_2.

Let $s_n = (x_1, x_2, \ldots, x_n)$ be a random sample realizing n identically distributed and independent random variables, having a common density function $f(X,\theta)$; and let S_n be the n-dimensional sample space. A decision function (d_n) is a measurable function on the sample space S_n, which assigns an action of A to each possible sample; where $A = \{a | a = a_1 \text{ or } a_2\}$. In other words, the decision of which action to take, given the n observations s_n is a function

$$d_n(X_1, X_2, \ldots, X_n) \in A.$$

Each decision function d_n can be represented by a partition of the n-dimensional sample space S_n into two disjoint sets $S_n^{(1)}$ and $S_n^{(2)}$, such that action a_1 is taken

if the sample point s_n falls in $S_n^{(1)}$ and a_2 is taken if s_n falls in $S_n^{(2)}$.

The sets w_1 and w_2 are associated with the hypothesis that θ is in w_1, and the alternative hypothesis that θ is in w_2. The action a_1 is called accepting the hypothesis H_1, and action a_2 is called rejection H_1. The function d_n, which when applied to the data leads to the accepting or rejecting of a hypothesis, is called a test of the hypothesis.

In the following presentation we shall be considering the following hypotheses:

$$\begin{aligned} H_0&:\theta = \theta_o \\ H_1&:\theta = \theta_1 > 0_o. \end{aligned} \tag{2.1}$$

Wald first worked with the simple hypotheses

$$\begin{aligned} H_0&:\theta = \theta_o \\ H_1&:\theta = \theta_1. \end{aligned}$$

He then generalized these simple hypotheses into (2.1), where H_1 became a composite hypothesis and where the desired properties were maintained.

To test these hypotheses sequentially, an observation is taken and a decision is made whether to accept H_0, accept H_1, or to take another observation. This process is repeated after each observation until one of the hypotheses is accepted. The probability is 1 that the sequential probability ratio test procedure will eventually terminate. Since the true value of θ is unknown, samples and statistical procedures

may lead the researcher to wrong conclusions.

If the true value of θ is equal to θ_0, but the test statistic leads to a decision of accepting H_1, an error has been committed. This error is called a Type I error and the probability of a Type I error is called an α-risk, and is denoted by α. Another possible error is the Type II error which is incurred when the true value of θ is equal to θ_1 but the test statistic indicates that H_0 is true. The probability of a Type II error is denoted by β. Power is defined as 1 minus the probability of a Type II error, and is the probability of rejecting H_0 whenever it is false.

Since the distribution of X is determined by a parameter θ, the probability of accepting H_0 is a function of θ. This function is called the operating characteristic function and is denoted by $L(\theta)$. Thus, for θ outside the region specified for it by H_0, the power of the test statistic is equal to $1 - L(\theta)$.

Reference is often made to the average (expected) sample number (A.S.N.). Because of the decision made after each observation, the number of observations, n, required by a sequential test is not predetermined, but does depend on the parameter θ.

In order to obtain a sequential procedure with the desired properties, Wald (16) introduced the sequential probability ratio test (S.P.R.T.). To derive a S.P.R.T. for hypotheses (1), it is assumed that θ_1 is an arbitrary

parameter greater than θ_0. Then $f(X,\theta_0)$ is the density function of X when H_0 is true, while $f(X,\theta_1)$ represents the distribution under H_1.

When H_0 is true, the likelihood (joint density) that a particular random sample $X_1, X_2, \ldots, X_m$ will be obtained is

$$P_{om} = f(X_1,\theta_0) \cdot f(X_2,\theta_0) \cdot \ldots \cdot f(X_m,\theta_0).$$

Similarly, when H_1 is true

$$P_{1m} = f(X_1,\theta_1) \cdot f(X_2,\theta_1) \cdot \ldots \cdot f(X_m,\theta_1).$$

Two positive constants A and B ($0<B<1<A<\infty$) are chosen. At each stage of the experiment, m, the probability ratio P_{1m}/P_{om} is computed. If, after taking m observations

$$B<\frac{P_{1m}}{P_{om}}<A, \tag{2.2}$$

the experiment is continued by taking one more observation. If

$$\frac{P_{1m}}{P_{om}} \geq A,$$

sampling is terminated with the rejection of H_o, and if

$$\frac{P_{1m}}{P_{om}} \leq B,$$

sampling is terminated with H_0 being accepted.

A convenient form of the above ratio is as follows:

$$\log \frac{P_{1m}}{P_{om}} = \log \frac{f(X_1,\theta_1)}{f(X_1,\theta_0)} + \ldots + \log \frac{f(X_m,\theta_1)}{f(X_m,\theta_0)}$$

$$= Z_1 + \ldots + Z_m.$$

At stage m of the experiment the cumulative sum $Z_1 + \ldots + Z_m$ is computed. If

$$\log B < Z_1 + \ldots + Z_m < \log A,$$

the experiment is continued by taking an additional observation. If

$$Z_1 + \ldots + Z_m \geq \log A$$

the process is terminated with the rejection of H_0. If

$$Z_1 + \ldots + Z_m \leq \log B$$

the process is terminated with the acceptance of H_0.

To determine A and B consider a sample $(X_1, \ldots, X_n)$ where H_0 is accepted in one case and where H_0 is rejected in the other. For H_0 to be accepted on the n the observation requires that

$$B < \frac{P_{1m}}{P_{om}} < A, \text{ for } m = 1, \ldots, n-1$$

and

$$\frac{P_{1n}}{P_{on}} \leq B,$$

but for H_0 to be rejected requires

$$B < \frac{P_{1m}}{P_{om}} < A, \text{ for } m = 1, \ldots, n-1$$

and

$$\frac{P_{1m}}{P_{om}} \geq A.$$

Wald stated that

$$A \leq \frac{1-\beta}{\alpha} \text{ and that } B \geq \frac{\beta}{1-\alpha}.$$

This can be proven by partitioning S_n into three sets such that $S_n = A_n \cup R_n \cup C_n$, where

$$A_n = \left\{(X_1, \ldots, X_n): \frac{f_1(X_i)}{f_o(X_i)} \leq B\right\}$$

$$R_n = \left\{(X_1, \ldots, X_n): \frac{f_1(X_i)}{f_o(X_i)} \geq A\right\}$$

$$C_n = \left\{(X_1, \ldots, X_n): B < \frac{f_1(X_i)}{f_o(X_i)} < A\right\}.$$

Set $\int_{A_n} \prod f(X_i) d\mu(X_i) = \begin{cases} \int_{A_n} \prod f(X_i) dX_i & \text{for the continuous case} \\ \int_{A_n} \prod f(X_i) & \text{for the discrete case.} \end{cases}$

$$P_\theta\left[\text{accepting } H_o\right] = \sum_{n=1}^{\infty} \int_{A_n} \prod_{i=1}^{n} f_0(X_i) d\mu(X_i)$$

for $\theta = 0, 1$. Therefore,

$$\beta = P_1\left[\text{accepting } H_0 \mid H_1 \text{ is true}\right]$$

$$= \sum_{n=1}^{\infty} \int_{A_n} \prod_{i=1}^{n} f_1(X_i)\, d\mu(X_i)$$

$$\leq B \sum_{n=1}^{\infty} \int_{A_n} \prod_{i=1}^{n} f_0(X_i)\, d\mu(X_i)$$

$$= B \cdot P\left[\text{accepting } H_0 \mid H_0 \text{ is true}\right]$$

$$= B(1-\alpha).$$

Therefore,

$$\frac{\beta}{1-\alpha} \leq B.$$

Similarly,

$$1-\beta = P\left[\text{rejecting } H_0 \mid H_1 \text{ is true}\right]$$

$$= \sum_{n=1}^{\infty} \int_{R_n} \prod_{i=1}^{n} f_1(X_1)\, d\mu(X_i)$$

$$\geq A \sum_{n=1} \int_{R_n} \prod_{i=1}^{n} f_0(X_i)\, d\mu(X_i)$$

$$= A \cdot P\left[\text{rejecting } H_0 \mid H_0 \text{ is true}\right]$$

$$= A\alpha,$$

or

$$A \leq \frac{1-\beta}{\alpha}.$$

Wald suggested that for practical purposes, set $A' = (1 - \beta)/\alpha$ and $B' = \beta/(1 - \alpha)$. If these limits are used for A and B, we have that the actual risk probabilities α' and β' satisfy:

$$\alpha' + \beta' \leq \alpha + \beta .$$

In order to substantiate the statement that the sequential procedure terminates with probability 1 the following lemma is given.

Lemma: The S.P.R.T. terminates with probability 1, for every $0 < B < 1 < A < \infty$ and for every θ.

Proof:

Let

$$Z_i = \log \frac{f_1(X_i)}{f_0(X_i)} , \quad (i = 1, 2, \ldots).$$

Since $X_1, X_2, \ldots,$ are independent random variables and identically distributed, so are $Z_1, Z_2, \ldots$.
Assume that for each θ,

$$P_\theta [Z = 0] < 1.$$

Without loss of generality assume that $0 < \sigma_\theta^2 = \mathrm{Var}_\theta (Z) < \infty$. If $\sigma_\theta^2 = 0$, this implies that $P_\theta [Z = C] = 1$ for some $C \neq 0$, and then obviously $S_n = \sum_{i=1}^{n} Z_i$ will eventually violate the inequality $-b < S_n < a$, where $-b = \log B$ and $a = \log A$. The assumption that $\sigma_0^2 > 0$ implies

that for sufficiently large r (constant integer), which depends on γ $(0 < \gamma < 1)$,

$P_\theta\left[|S_r| > a + b\right] \geq 1 - \gamma$. This can easily be shown by using the central limit theorem for $(S_r - r\mu_\theta) / \sqrt{r\,\sigma_\theta^2}$, with $\mu_\theta = E_\theta[Z]$.

Finally, if N designates the sample size in the S.P.R.T.,

$$P_\theta[N < \infty] = 1 - \lim_{p \to \infty} P_\theta[N \geq pr].$$

But

$$P_\theta[N > pr] = P_\theta[-b < S_n < a \text{ for all } n = 1,2, \ldots, pr]$$

$$\leq P_\theta[-b < S_r < a,\ |S_{2r} - S_r| < a + b, \ldots,$$

$$|S_{pr} - S_{(p-1)r}| < a + b]$$

$$\leq (P_\theta[|S_r| < a + b])^{p-1} \leq \gamma^{p-1}.$$

Thus,

$$\lim_{p \to \infty} P_\theta[N > pr] \leq \lim_{p \to \infty} \gamma^{p-1} = 0.$$

A strength is associated with a given sequential test, which correspondends to α and β. Two sequential tests, S and S' are said to be of equal strength if the values of α and β of S are equal to the corresponding values α' and β' of S'. If $\alpha < \alpha'$ and $\beta \leq \beta'$, or if $\alpha \leq \alpha'$ and $\beta < \beta'$, S is stronger than S'. The strength of a given sequential test is denoted (α, β).

Hence, if a researcher has two sequential tests with the same strength, a logical question to consider is whether the expected number of observations (A.S.N.) are equal for desired range of θ. To derive a formula for the A.S.N. of a test, the operating characteristic function is needed.

$L(\theta)$ has been defined as the probability that the sequential process terminates with the acceptance of H_0 when θ is the true value of the parameter.

Consider the expression

$$\left(\frac{f(X,\theta_1)}{f(X,\theta_0)}\right)^{h(\theta)} \qquad \text{where } h(\theta) \neq 0 \text{ and}$$

$$\int_{-\infty}^{\infty}\left(\frac{f(X,\theta_1)}{f(X,\theta_\theta)}\right)^{h(\theta)} f(X,\theta)dX = 1 \tag{2.3}$$

in the continuous case or

$$\sum_{X}\left(\frac{f(X,\theta_1)}{f(X,\theta_o)}\right)^{h(\theta)} f(X,\theta) = 1 \tag{2.4}$$

in the discrete case. Wald proved that there was only one θ satisfying the above equations after some slight restriction were placed on $f(X,\theta)$.

Hence, for any given value θ, the function of X given by

$$f^*(X,\theta) = \left(\frac{f(X,\theta_1)}{f(X,\theta_0)}\right)^{h(\theta)} f(X,\theta) \tag{2.5}$$

is a probability density function.

Consider the following hypotheses:

H : $f(X,\theta)$ is the true distribution of X

H* : $f^*(X,\theta)$ is the true distribution of S.

Assume $h(\theta) > 0$ and consider the S.P.R.T. S* for testing H against H*. Continue sampling procedure as long as

$$B^{h(\theta)} < \frac{f^*(X_1,\theta) \ldots f^*(X_m,\theta)}{f(X_1,\theta) \ldots f(X_m,\theta)} < A^{h(\theta)} , \tag{2.6}$$

but accept H if

$$\frac{f^*(X_1,\theta) \ldots f^*(X_m,\theta)}{f(X_1,\theta) \ldots f(X_m,\theta)} \leq B^{h(\theta)} , \tag{2.7}$$

or reject H if

$$\frac{f^*(X_1,\theta) \ldots f^*(X_m,\theta)}{f(X_1,\theta) \ldots f(X_m,\theta)} \geq A^{h(\theta)} . \tag{2.8}$$

Since

$$\frac{f^*(X,\theta)}{f(X,\theta)} = \left(\frac{f(X,\theta_1)}{f(X,\theta_0)}\right)^{h(\theta)} ,$$

the inequalities (2.6), (2.7), and (2.8) are equivalent to

$$B < \frac{f(X_1,\theta_1) \ldots f(X_m\theta_1)}{f(X_1,\theta_0) \ldots f(X_m,\theta_0)} < A$$

$$\frac{f(X_1,\theta_1) \ldots f(X_m,\theta_1)}{f(X_1,\theta_0) \ldots f(X_m,\theta_0)} \leq B$$

$$\frac{f(X_1,\theta_1) \ldots f(X_m\theta_1)}{f(X_1,\theta_0) \ldots f(X_m,\theta_0)} \geq A \ .$$

But these inequalities are the same as the ones derived from S.P.R.T. S in (2.1). Thus, the probability of accepting H_0 when $\theta = \theta_0$ is the same as the probability that the test S* will lead to the acceptance of H when $f(X,\theta)$ is the true distribution of X.

Let the strength of S* be (α', β'). It follows that $A^{h(\theta)} \cong \frac{1-\beta'}{\alpha'}$ and $B^{h(\theta)} \cong \frac{\beta'}{1-\alpha'}$.

Hence α' is approximately equal to

$$\frac{1 - B^{h(\theta)}}{A^{h(\theta)} - B^{h(\theta)}}$$

Since $\alpha' = 1 - L(\theta)$,

$$L(\theta) \quad \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}} \ . \tag{2.9}$$

The same result is obtained if $h(\theta) < 0$; and; therefore, (2.9) gives an approximation to the operating characteristic function.

Using equation (2.9), one is able to find the A.S.N. function of a S.P.R.T.

Let n denote the number of observations required by the procedure, and let $E_\theta(n)$ be the expected value of n when θ is the true value of the parameter. The function $E_\theta(n)$ is the A.S.N. function, neglecting the excess of

P_{1m} / P_{om} over the termination bounderies.

Let N be an integer sufficiently large to allow the probability that $n \geq N$ to be neglected.

Then

$$Z_1 + \ldots + Z_N = (Z_1 + \ldots + Z_n) + (Z_{n+1} + \ldots + Z_N) \tag{2.10}$$

where

$$Z_i = \log \frac{f(X_i,\theta_1)}{f(X_i,\theta_0)} .$$

Upon taking the expected value of both sides of (2.10), it follows that

$$NE(Z) = E(Z_1 + \ldots + Z_n) + E(Z_{n+1} + \ldots + Z_N) \tag{2.11}$$

where $Z = \log \frac{f(X,\theta_1)}{f(X,\theta_0)}$.

Since for $i > n$, the random variable Z_i is distributed independently of n, the following is obtained:

$$E(Z_{n+1} + \ldots + Z_N) = E\left[E(Z_{n+1} + \ldots + Z_N) \mid n\right]$$

which

$$= E\left[(N - n)\, E(Z)\right] \tag{2.12}$$

or

$$= E(N - n)\, E(Z)$$

and finally

$$E(Z_{n+1} + \ldots + Z_N) = NE(Z) - E(n)E(Z). \tag{2.13}$$

From (2.11) and (2.13) we obtain

$$E(Z_1 + \ldots + Z_n) - E(n)E(Z) = 0.$$

Hence

$$E(n) = \frac{E(Z_1 + \ldots + Z_n)}{E(Z)}, \text{ for } E(Z) \neq 0. \tag{2.14}$$

If θ is the true value of the parameter, then by definition $E_\theta(n) = E(n)$. Neglecting excess at the boundaries, the random variables $(Z_1 + \ldots + Z_n)$ can assume the values log A and log B with probabilities $1 - L(\theta)$ and $L(\theta)$, respectively, which implies $E(Z_1 + \ldots + Z_n) \cong [1 - L(\theta)] \log A + L(\theta) \log B$. Hence,

$$E_\theta(n) \quad \frac{[1 - L(\theta)] \log A + L(\theta) \log B}{E_\theta(Z)}. \tag{2.15}$$

While it was proved that the probability is 1 that the sequential probability ratio test will eventually terminate; there is a possibility of an unexpected long series of observations. Armitage considered truncated sequential designs in which the Wald S.P.R.T. was modified so that the power and α-risk remained nearly the same. Armitage chose to call these truncated designs, closed designs. Open and closed procedures under different probability distributions will be discussed in the subsequent sections.

3. The S.P.R.T. in the Binomial Case

The basic use of the sequential binomial test is that of choosing the better of two treatments. This is done by forming a sequence of preferences. These preferences can be made subjectively by the patients or can be more objective.

In the two-tailed test a strength of $(2\alpha, \beta)$ will be discussed.

Let θ be the probability that the preference is treatment A. The hypotheses under question are:

$$\begin{aligned} H_0&: \theta = \theta_0 = \tfrac{1}{2} \\ H_1&: \theta = \theta_1 < \tfrac{1}{2} \\ H_2&: \theta = \theta_2 > \tfrac{1}{2} . \end{aligned} \tag{3.1}$$

If A is really better than some treatment B, then $\theta > \frac{1}{2}$. If B is better, then $\theta < \frac{1}{2}$. For simplicity it was assumed that θ_1 and θ_2 are symmetrical about the value $\frac{1}{2}$. To derive the test procedure for the above hypotheses, only H_0 and H_1 will be considered with strength (α, β). Then because of the symmetry assumed between θ_1 and θ_2, the test will be extended to include the hypotheses (3.1) and have an overall strength approximately equal to $(2\alpha, \beta)$.

Let X_i denote the outcome of the ith pair; i.e., $X_i = 1$ if the ith pair shows a preference for treatment A, and $X_i = 0$ if the preference was for B. For the first m pairs observed, the probability of the observed sample $(X_1, \ldots, X_m)$ in that order is equal to

$$\theta^{d_m} (1 - \theta)^{m-d_m} \tag{3.2}$$

where d_m denotes the number favoring treatment A in the first m units.

Under H_1, the likelihood of (3.2) becomes

$$P_{1m} = \theta_1^{d_m} (1 - \theta_1)^{m-d_m}$$

and under H_o (3.2) becomes

$$P_{om} = \theta_o^{d_m} (1 - \theta_o)^{m-d_m} .$$

At each stage of the experiment, denoted generally by m, the logarithm of the ratio $\frac{P_{1m}}{P_{om}}$ is computed:

$$\log \frac{P_{1m}}{P_{om}} = (d_m) \log \frac{\theta_1}{\theta_o} + (m - d_m) \log \frac{1 - \theta_1}{1 - \theta_o} . \tag{3.3}$$

Preference testing is continued as long as

$$\log B = \log \frac{\beta}{1 - \alpha} < \log \frac{P_{1m}}{P_{om}} < \log \frac{1 - \beta}{\alpha} = \log A.$$

If $\log \frac{P_{1m}}{P_{om}} \geq \log \frac{1 - \beta}{\alpha}$, H_o is rejected; but if (3.4)

$$\log \frac{P_{1m}}{P_{om}} \leq \log \frac{\beta}{1 - \alpha}, \quad H_o \text{ is accepted.} \tag{3.5}$$

By substituting (3.3) into (3.4) and (3.5) and solving for d_m one obtains critical values for each step m corresponding to given values of θ_o and θ_1. For rejecting and

accepting H_o, respectively, the following inequalities are given:

$$d_m \geq \frac{\log \frac{1-\beta}{\alpha}}{\log \frac{\theta_1}{\theta_o} - \log \frac{1-\theta_1}{1-\theta_o}} + m \frac{\log \frac{1-\theta_o}{1-\theta_1}}{\log \frac{\theta_1}{\theta_o} - \log \frac{1-\theta_1}{1-\theta_o}}, \quad (3.6)$$

$$d_m \leq \frac{\log \frac{\beta}{1-\alpha}}{\log \frac{\theta_1}{\theta_o} - \log \frac{1-\theta_1}{1-\theta_o}} + m \frac{\log \frac{1-\theta_o}{1-\theta_1}}{\log \frac{\theta_1}{\theta_o} - \log \frac{1-\theta_1}{1-\theta_o}}. \quad (3.7)$$

Armitage (3) gives tabular values for given θ_1's and θ_2's for the researcher wanting critical values at each stage of the experiment. A common procedure is to plot (3.6) and (3.7) on cordinate graph paper.

Similarly, by considering only H_o and H_2 in (3.1) another set of critical regions can be established. These two sets of critical regions are symmetric with respect to the horizontal axis. For this two-sided case X_i take the values 1 and -1 (instead of 1 and 0) in the preferences for A and B, respectively.

By letting X_i take values of -1 and +1, Armitage overcame two complications which arise from (3.1) if X_i assumed only 0 and +1. A sample path taking values of +1 and -1 contradicts corresponding one-sided tests only if it crosses

both dotted lines in figure 1 (see Sobel and Wald, Ann. Math. Stat., 1949, for details). Armitage stated that sampling was terminated with the acceptance of H_o if the sample path crossed both dotted lines because a sample path of this kind would have terminated in the same conclusion if considered in either of the two one-sided tests.

A typical open design is given in fig. 1, where

$$-a_1 = \frac{\log \frac{1-\beta}{\alpha}}{\log \frac{\theta_1}{\theta_o} - \log \frac{1-\theta_1}{1-\theta_o}} \tag{3.8}$$

$$a_2 = \frac{\log \frac{\beta}{1-\alpha}}{\log \frac{\theta_1}{\theta_o} - \log \frac{1-\theta_1}{1-\theta_o}} \tag{3.9}$$

and

$$-b = \frac{\log \frac{1-\theta_o}{1-\theta_1}}{\log \frac{\theta_1}{\theta_o} - \log \frac{1-\theta_1}{1-\theta_o}} .$$

The equations of the outer boundaries (indicating significant differences between treatments) are:

U: $y = a_1 + bm$

L: $y = -a_1 - bm.$

The equations of the inner boundaries (which close the trial with no differences established) are:

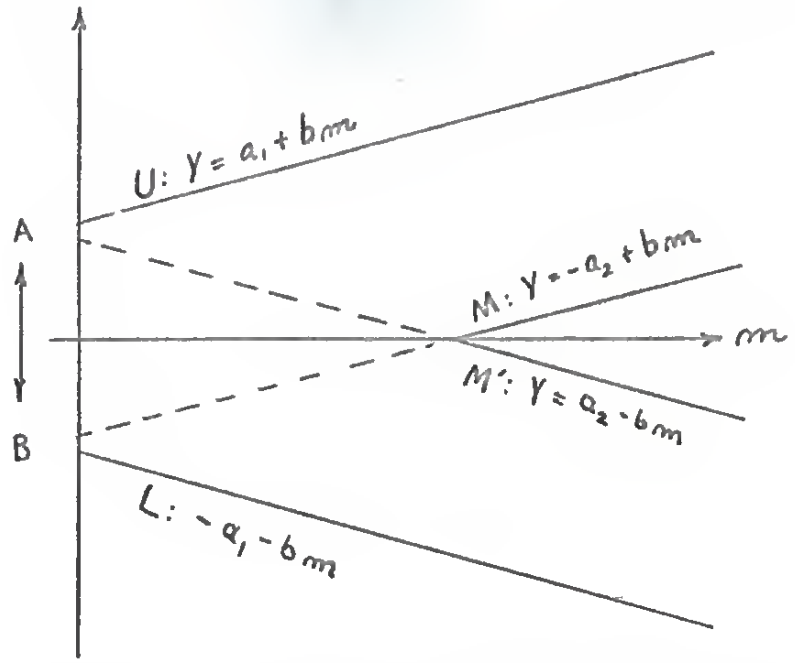


Fig. 1. SCHEMATIC REPRESENTATION OF AN OPEN DESIGN.

M: $y = -a_2 + bm$

M': $y = a_2 - bm$,

where in both sets of equations the a's are intercepts and the b's are the slopes of the lines.

The testing can be carried out on ordinary graph paper by plotting the boundaries as given in Fig. 1. As results begin to become available each preference for A is plotted on the graph by moving one unit horizontally to the right and one unit vertically up, while a preference for B is noted by one sliding one unit to the right and then one unit vertically down. If no preference is given (a tie), nothing is recorded. This process is continued until one of the boundaries is reached.

A common method of assessing the relative efficiency of two therapeutic procedures is to form 2 groups of patients treating each group with a different therapy, and compare the favorable responses toward one treatment in each group. Up to this point tied observations were ignored completely in the analysis by assuming that a preference had to be made for one of the treatments or that the pair would be ignored if a tie occurred. We shall now consider the effects of ties on hypotheses (3.1). Armitage (3) indicated that the following procedure for comparison of two proportions could be used on observations that are naturally paired as well as those which are not paired according to some criterion.

As patients entered the trial, they were formed into pairs with each randomly receiving either treatment A or B. Each pair of subjects could result in one of four possible outcomes: (S,S), (F,F), (S,F), or (F,S) where the first of each of these ordered pair represented the success (S) or failure (F) of treatment A and the second represented the outcome of B. The two pairs (S,S) and (F,F) were tied pairs; while (S,F) and (F,S) showed a preference for A and B, respectively.

Thus, a sequence of preferences had been formed, which follows the discussion already presented but which also had some interesting properties.

Suppose that the true probability of success for A is π_1 while the probability of success for B is π_2. Then on

the assumption of random pairing, probabilities were figured for each of the four ordered pairs:

POSSIBLE OUTCOME	PROBABILITY
(S,S)	$\pi_1 \pi_2$
(F,F)	$(1-\pi_1)(1-\pi_2)$
(S,F)	$\pi_1 (1-\pi_2)$
(F,S)	$(1-\pi_1) \pi_2$

Hence, the probability of an untied pair, denoted by , is given by the formula,

$$\phi = \pi_1(1 - \pi_2) + (1 - \pi_1)\pi_2.$$

Considering only untied pairs, the proportion yielding A preferences, denoted by θ, is

$$\theta = \frac{\pi_1(1 - \pi_2)}{\phi} = \frac{\pi_1(1 - \pi_2)}{\pi_1(1 - \pi_2) + (1 - \pi_1)\pi_2},$$

and the proportion yielding B preferences is

$$1 - \theta = \frac{(1 - \pi_1)\pi_2}{\phi} = \frac{(1 - \pi_1)\pi_2}{\pi_1(1 - \pi_2) + (1 - \pi_1)\pi_2}.$$

The series of preferences is therefore a binomial sequence with the probability of an A preference equal to θ. Table 4.1 in Armitage (3) gives values of ϕ and θ for various values of π_1 and π_2.

Some of the main points to notice are: (a) When $\pi_1 = \pi_2$, $\theta = \frac{1}{2}$; (b) when π_1 is greater than π_2, $\theta > \frac{1}{2}$; and (c) when

π_1 is less than π_2, $\theta < \frac{1}{2}$. It is now clear that the same hypotheses apply. If the upper boundary is reached, we have good evidence that $\theta > \frac{1}{2}$, which implies $\pi_1 > \pi_2$.

Using the tables given for restricted binomial procedures, one obtains a value N for the number of untied subjects needed for the specified risks. By taking N and dividing it by the probability of untied observations ϕ, one obtains a new sample number larger than N, suggesting how large the actual sample size will need be (See section 4).

With data paired naturally (equivalently stratified or matched) the value ϕ for untied pairs is always less than $\pi_1(1 - \pi_2) + (1 - \pi_1)\pi_2$, which is valid for random pairing. Armitage (3) indicated that this is not a disadvantage, since for given values of π_1 and π_2, stratification will tend to give a value of θ further away from $\frac{1}{2}$ than would be expected from the formula for θ. This means that one can achieve higher power of detecting a particular difference in π_1 and π_2 than could be obtained in random pairing. This reduces the average number of preferences required.

4. Truncated and Restricted S.P.R.T.

Wald (16) indicated that truncation could be considered in large sampling without causing much change in power or risk.

In 1952 Bross (6) introduced the idea of a closed design for clinical trials. Bross constructed two designs which were believed to have been constructed mostly by trial and error. This was the first attempt to build a truncated open design to control the risks.

In 1957 Armitage (2) introduced his restricted procedures which are similar to Bross' but which were less restrictive to the researcher. The important feature of Armitage's procedure is in the truncation of the Wald's S.P.R.T. At each stage a decision is made either to accept one of the hypotheses or to continue sampling, but one of the hypotheses is chosen by the time the Nth observation is recorded. The truncation of Wald's S.P.R.T. at a point N may change the strength of the untruncated sequential test. Stockman and Armitage (15) investigated the effect of truncation on the strength, and discussed an exact method for finding the total number of paths that a sample path may take in a restricted binomial procedure. Armitage (2) then offered an approximation by studying a corresponding diffusion process.

Consider Fig. 2 for the exact method. The bounds corresponded to the lines $y = a_1 + bm$ and $y = -a_2 + bm$ given in Fig. 1. The area within the boundaries is the continuation

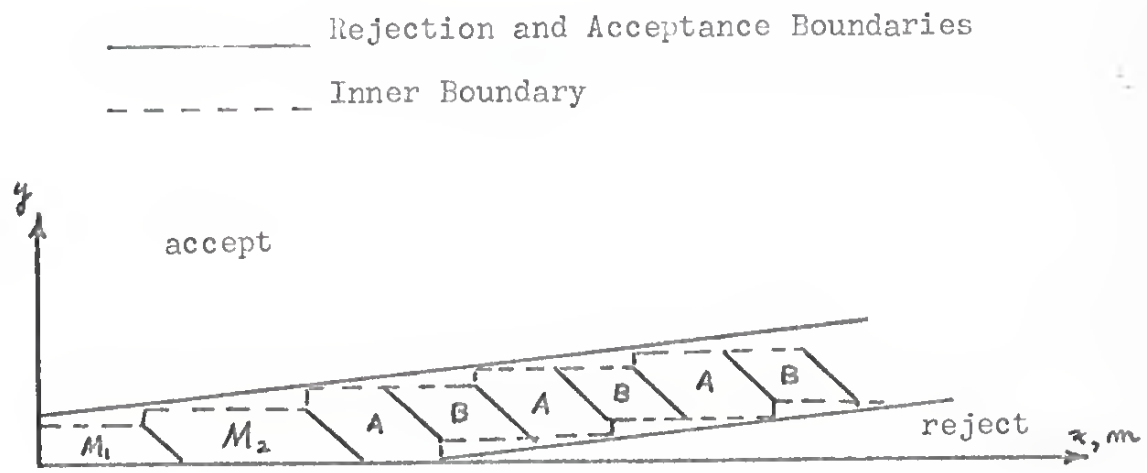


Fig. 2. STOCKMAN AND ARMITAGE'S EXACT METHOD.

region. Each of the lattice diagrams (M_1, M_2, A, B) had the following properties: 1) In a random sample, the probability of reaching the point $(x, y) = N\,\theta^y\,(1 - \theta)^x$, where N was the number of paths to (x, y) which were not interrupted by the acceptance or rejection boundaries. 2) The number of paths from (x_1, y_1) to $(x_1 + x, y_1 + y)$, if no paths were interrupted by the boundaries, was equal to the number of ways of chosing the positions of the x unit displacements parallel to horizontal axis x in a total of (x + y) displacements, which is:

$$\binom{x + y}{x} .$$

With each lattice diagram a matrix was constructed such that the (i, j)th element represented the number of paths from the ith point of the left hand diagonal side to the jth point of

the right hand diagonal side (counting from the top), all of the elements being binomial coefficients. Further, Stockman and Armitage indicated that the martices could be multiplied together so that the (i, j)th element of the product of 2 adjacent matrices, $S = (s_{ij})$ and $T = (t_{ij})$, was the number of paths from the ith point of the left hand side of S to the jth point of the right hand side of T.

Having noted that the lattice diagrams A and B in Fig. 2, repeat after M_2, the authors stated that the total number of paths from the origin to the points of a particular lattice diagram could be expressed generally by $G = M_1M_2\ (AB)^n$ (the matrices represent their corresponding diagram), assuming that the process terminated with the right hand side of the nth-B diagram. It was then pointed out that the last element of the row matric G represented the number of admissible paths crossing the lower boundary for the first time. By multiplying this number by the probability of reaching this point $(\theta^y(1 - \theta)^x)$, the reserachers found the probability of a sample path crossing the acceptance region for the first time after a specified number of observations. By setting $\theta = \theta_o$, the value specified under the null hypothesis, one is able to estimate α for various n in the expression G. By setting $\theta = \theta_1$, the probability of accepting H_0 when $\theta = \theta_1$ under H_1, β can be calculated for the one sided test for various n.

It is readily apparent that the above presentation is interesting from a theoretical point of view, but practical

applications would be very tedious.

In an attempt to overcome some of the shortcomings of Stockman and Armitage's exact method, Armitage (2) introduced a diffusion approximation. Admitting that the process had not been fully investigated, the researcher compared the diffusion method with the exact method for fixed α and β. The diffusion process proved to be fairly satisfactory in the specific cases compared.

Armitage stated that the discrete steps of m could be replaced by continuous movement in time, where the time unit corresponded to a single observation. He claimed that the random variable could be approximated by the one-dimensional diffusion process with drift, variance and an absorbing barrier.

The upper boundary of the open binomial process, U: $y = a_1 + bm$ and the values for a_1 and b given in (3.8) and (3.10) was maintained. The lower bound M was replaced by a vertical boundary M' with equation $m = N$. Then N became the maximum number of observation to be taken in this closed procedure. The next step was to determine N to maintain prescribed α and β-risks. Using

$$y_m = \sum_{i=1}^{m} X_i$$

where the X_i are independent variates taking the values +1 or -1 with probabilities θ and $1 - \theta$ respectively, Armitage

indicated that the random variable y_m could be approximated by a diffusion process with drift $(\mu - b)$ per time unit, growth in variance at a rate σ^2, and an absorbing barrier at a, where

$$\mu = 2\theta - 1 \text{ and } \sigma^2 = 4\theta (1 - \theta) .$$

To ensure the procedure's power $(1 - \beta)$ the author, using a result of Bartlett (5), set

$$\beta = F \left(\frac{a_1 - m_1 N}{\sigma_1 \sqrt{N}}\right) - \exp \left(\frac{2a_1 m_1}{\sigma_1^2}\right) F \left(\frac{-a_1 - m_1 N}{\sigma_1 \sqrt{N}}\right) ,$$

where a_1 is defined through H_1 by (3.8),

$$\sigma_1^2 = 4\theta_1 (1 - \theta_1) ,$$

$$m_1 = 2\theta_1 - 1 - \frac{2 \ln \left(\frac{1}{\sigma_1}\right)}{\ln \frac{\theta_1}{1 - \theta_1}} ,$$

and

$$F(u) = \int_{-\infty}^{u} (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}t^2)dt.$$

Finally, M' was replaced by a wedge-shaped boundary M", and new boundaries analogously under H_2 were formed, making a two sided test. This is illustrated in Fig. 3.

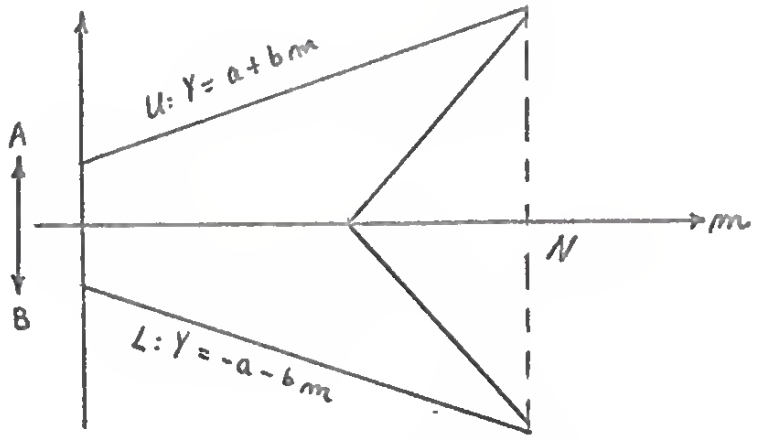


Fig. 3. SCHEMATIC REPRESENTATION OF A RESTRICTED DESIGN.

It should be noted that any sample path reaching the wedge formed by M" causes an aceptance of H_o; that the bound M" is drawn at an angle of 45°, and that a sample path crossing M" cannot cross either U or L. Hence the replacement of M' by M" does not affect the probability of reaching any boundary point on U or L, but does reduce the average sample number.

Values for these boundaries are given in Armitage (3). Wetherill (17) indicated that this boundary probably could be improved but did not make any suggestions.

It should be noted that the middle boundary M" in the restricted procedure helps delete the expected or average number of subjects required, but that this average number will be greater than the corresponding number given for the open design, since the sample path may cross the middle boundary of the open plan much sooner than in the restricted procedure.

In Armitage (3) proposed what he called a skew design. This design follows intuitively from the binomial restricted design. If one was interested in detecting the superiority of a treatment A over B, where A is a new therapy and B is a standard, the researcher might not be interested in how bad B is; and thus, he could use a skew design (see Fig. 4).

Examples for both the restricted and skewed binomial designs are given in section 5.

Choi (7) introduced a truncated sequential design for the random binomial sequence using a fundamental equation of Markov chains. He proposed the same hypotheses as were given in (3.1).

Choi defined "path points at n" as the points which can be reached at the nth sampling step.

Let $S_o(\theta) = s_i$ be a vector such that the ith component s_i is the probability of reaching the ith path point at $n = c$ when θ is the probability of a preference for A.

Let $T_1(\theta)$ be the transition matrix from path points at $n = c$ to those at $n = c + 1$. $T_1(\theta)$ is a $(c + 1)X(c + 1)$ matrix constructed by adding a superfluous column of zeroes to a $(c + 1)X\ c$ matrix.

Let $T_2(\theta)$ be the $(c + 1)X(c + 1)$ transition matrix from path points at $n = c$ to those at $n = c + 2$.

Figure 5 is presented as an aid.

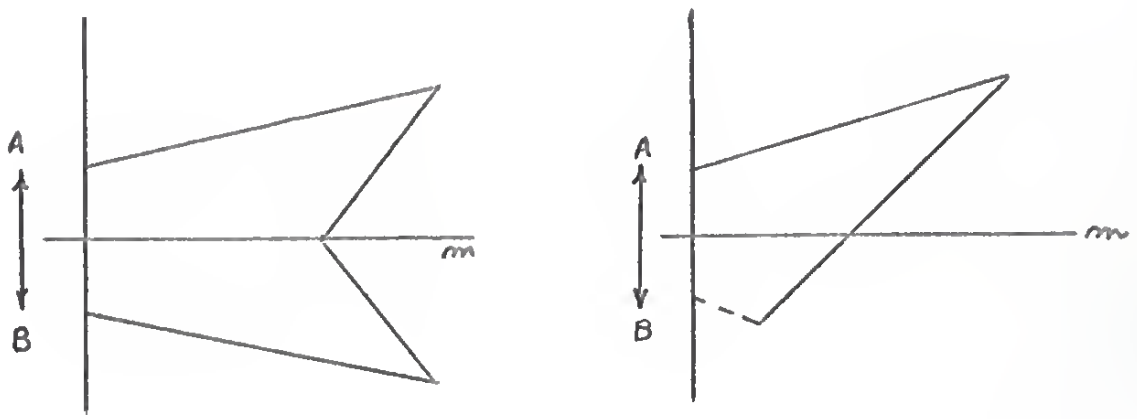


Fig. 4. THE METHOD OF FORMING SKEW RESTRICTED DESIGNS.

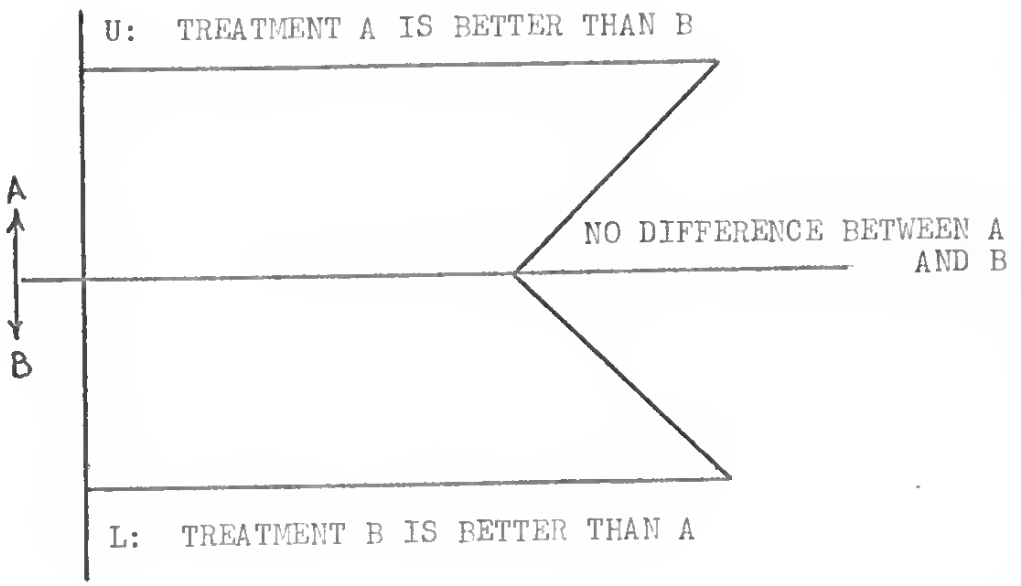


Fig. 5. TRUNCATED SEQUENTIAL CLINICAL TRIALS.

It follows that

$$S_{o(p)} = \left[p^c,\ \binom{c}{1} p^{c-1} q,\ \binom{c}{2} p^{c-2} q^2,\ \ldots,\ q^c\right],$$

$$T_2(p) = \begin{bmatrix} o & o & . & & . & . & . & o \\ p^2 & 2pq & q^2 & & o & . & . & o \\ o & p^2 & 2pq & & q^2 & . & . & o \\ & & & \ldots & & & & \\ o & o & . & . & . & p^2 & 2pq & q^2 \\ o & o & . & . & . & & & o \end{bmatrix},$$

and

$$T_1(p) = \begin{bmatrix} o & o & . & . & . & . & . & o \\ p & q & o & . & . & . & . & o \\ o & p & q & o & . & . & . & o \\ & & & \ldots & & & & \\ o & o & & & . & p & q & o \\ o & o & \ldots & & & & & o \end{bmatrix}$$

It was noted that the upper or lower boundaries could be reached only at $n = c + 2_j$, $j = 0, 1, 2, \ldots, h$. (It was assumed that $n_t = c + 2h + \delta$ where $\delta = 0$ or 1 depending on truncation.)

Letting $S_i(p)$ be a vector where the ith component is the probability of reaching the ith path point at $n = c + 2j$, the author used a fundamental equation of Markov chains:

$$S_j(\theta) = S_o(\theta)\ T_2^{\,j}(\theta),\ j = 0, 1, \ldots, h,$$

where $T_2^{\circ}(\theta)$ was defined as the $(c + 1)X(c + 1)$ identity matrix.

Let $S'_{\delta}(\theta)$ denote a vector such that the ith component is the probability of terminating at the ith path point when $m = n_t$. Chio stated that

$$S'(\theta) = \begin{cases} S_o(\theta)T_2^{h}(\theta) & \text{if } \delta = 0 \\ S_o(\theta)T_2^{h}(\theta)T_1(\theta) & \text{if } \delta = 1 . \end{cases}$$

By defining V_j, $j = 1, 2, \ldots, c$, as follows:

$$V_1 = [1, 0, \ldots, 0]'$$
$$V_2 = [0, 1, \ldots, 0]'$$
$$\ldots$$
$$V_c = [0, \ldots, 1, 0]' ,$$

he found exact probabilities of accepting H_o, H_1, and H_2 when the probability of an A preference was 0. Denoting these probabilities by $\alpha_o(\theta)$, $\alpha_1(\theta)$, and $\alpha_2(\theta)$; the author gave the following equations:

$$\alpha_2(\theta) = \sum_{j=o}^{h} S_j(\theta)V_1$$

$$\alpha_o(\theta) = S_{\delta}'(\theta) \sum_{j=2}^{c} {}_{-\delta} V_j$$

$$\alpha_1(\theta) = \sum_{j=o}^{h} S_j(1 - \theta)\, V_1 .$$

It should be noted that the wedged boundary in Fig. 5 was used for the same reasons that Armitage used M" in Fig. 3; and that the above method fails to offer a quick decision as is obtained in Armitage's restricted binomial procedure when the difference between two treatments is large.

5. Some Applications in Medical Trials

Example 1. Investigating the effects of hydrocortisone hemisuccinate as an inhalent for children with asthma, Smith (14) preformed a double blind trial, with a placebo preparation indistinguishable from the hydrocortisone and unidentified by the investigators until after the end of the trial. Children were randomly placed in two treatment groups, and each child used the specified inhalant daily for a month. The value of the treatment was determined by respiratory tests, by clinical records, and by collation of the different types of evidence. It was then judged as being a success or failure. Before the trial began it was estimated that 15% of the children would probably benefit from the placebo, and that an increase of 50% would be sufficient to warrant notice. Setting $\pi_1 = .65$ and $\pi_2 = .15$ gives $\theta = .913$. The open designs discussed in section 3 with $\theta_1 = .90$, $2\alpha = .05$, $1 - \beta = .95$ was used. The middle boundary was crossed at the 6th pair. The trial was continued a little longer, by which time 10 untied pairs had been observed and there were 4 successes out of 28 children treated with the placebo and 6 successes out of 29 children treated with hydrocortisone.

Example 2. Using the restricted design Robertson and Armitage (10) compared 2 hypotensive agents, phenactropinium chloride and trimetaphen between patients. The test criterion was the length of time required for the systolic blood pressure

to recover to a level of 100 mm, Hg, after it had been lowered during operation by one of the two drugs. The preference was determined by noting which of the 2 recovery times in each pair was shorter.

The authors specified $2\alpha = .05$, $1-\beta = .95$, $\theta_1 = \theta_2 = .75$, and $N = 62$. Since the drugs were used widely, results flowed in fairly quickly, and it was felt worthwhile to use a rather large design to maintain a greater power. The path reached the middle boundary at the 49th preference, after which one more pair of patients was treated. Three patients had equal recovery times and thus, provided no preference. In all 53 pairs of patients were used in the trial. It is interesting to note that if this trial had not been run sequential, a preference for trimetaphen might have been signified after the first 12 preferences.

Example 3. Using a restricted design with $2\alpha = .05$, $1-\beta = .95$, $\theta_2 = \theta_3 = .09$, and $N = 19$, Marshall and Shaw (8) used a skew design to test anticoagulants in the treatment of cerebral infarction. The preferences were "untied" pairs of patients differing in their survival experience 6 weeks after the beginning of the treatments. When 8 preferences had been obtained, 5 were for the control treatment without anticoagulants and 3 were for the anticoagulants. At this time the path had entered the pointed area from which the outer boundary favoring anticoagulants was inaccessible. It was regarded unethical to continue this investigation to see

whether the anticoagulant treatment could be shown to be worse than the standard treatment.

6. Continuous Measurements with Unknown Variability

Commonly an experimenter measuring quantitative responses will have some knowledge of the variation of a response, but he is not willing to rely on this estimate to use a sequential procedure assuming σ^2 known. Thus, in the case of independent, normally distributed random variables, if σ^2 is unknown, and we wish to test two-sided hypotheses concerning the mean μ, we apply a sequential t-test. In the present section we describe the sequential t-test. In the special case of pairing observations, let X_i be the difference of two random variables X_{1i} and X_{2i} representing responses from treatments A and B, respectively.

In Armitage (3) suggested the following set of hypotheses:

$$H_o: \frac{\mu}{\sigma} = 0 \qquad (6.1)$$

$$H_1: \left|\frac{\mu}{\sigma}\right| = \delta \ .$$

It was noted that in the more general case one would be testing whether $\mu = \mu_o$ against the alternative that $\mu = \mu_o + \delta\sigma$ or $\mu = \mu_o - \delta\sigma$, which is reducible to (6.1) by subtracting μ_o from all the observations.

Ruston (11) considered the same hypotheses but stated H_1 as $\mu = K\delta\sigma$ where K is the unspecied sign of μ (K = -1 or +1).

Now, the likelihood function of a set of n independent observations X_i, $(i = 1, \ldots, n)$, identically distributed as $N(K\delta\sigma, \sigma^2)$ is

$$f(X_1, \ldots, X_n; \delta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - K\delta\sigma)^2\right\}$$

which is equal to

$$f(X_1, \ldots, X_n; \delta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)S^2 + (K|t|S - \sqrt{n}\,\delta K\sigma^2)\right]\right\}. \qquad (6.2)$$

where

$$(n-1)S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 \text{ and } t^2 = \frac{n\bar{X}^2}{S^2}.$$

Since the density function (6.2) depends only on the sufficient statistics $(S^2, K|t|)$, and the distribution of S^2 does not depend on δ, we can let the likelihood function be represented by the density function of the non-central t, with $n-1$ d.f., $f(t^2|\delta)$. This density function is given by:

$$\frac{2(n-1)^{\frac{1}{2}(n-1)}}{B(\frac{1}{2}(n-1), \frac{1}{2})(n-1+t^2)^{\frac{1}{2}n}} M\left(\tfrac{1}{2}n, \tfrac{1}{2}, \frac{n\delta^2t^2}{2(n-1+t^2)}\right) e^{-\frac{1}{2}n\delta^2} \qquad (6.3)$$

$B(\alpha,\beta)$ represents the beta function, and M is the confluent hypergeometric function:

$$M(\alpha,\gamma,X) = \sum_{i=0}^{\infty} \frac{\Gamma(\gamma)\,\Gamma(\alpha+i)}{\Gamma(\alpha)\,\Gamma(\gamma+i)} \frac{X^i}{i!} . \tag{6.4}$$

Hence, the likelihood ratio becomes

$$\lambda_n = \frac{f(t^2|\delta)}{f(t^2|0)} = M(\tfrac{1}{2}n,\ \tfrac{1}{2},\ \frac{n\,\delta^2 t^2}{2(n-1+t^2)})\ e^{-\frac{1}{2}n\delta^2} . \tag{6.5}$$

Substituting,

$$u^2 = \frac{nt^2}{n-1+t^2} = \frac{\left[\sum_{i=1}^{n} X_i\right]^2}{\sum_{i=1}^{n} X_i^2} , \tag{6.6}$$

we can write the likelihood ratio in the form:

$$\lambda_n = e^{-\frac{1}{2}n\delta^2}\ M(\tfrac{1}{2}n,\ \tfrac{1}{2},\ \tfrac{1}{2}\delta^2 u^2). \tag{6.7}$$

As before, the sequential procedure consist of taking additional observation as long as

$$\frac{\beta}{1-\alpha} < \lambda_n < \frac{1-\beta}{\alpha} . \tag{6.8}$$

Thus, H_o is accepted as soon as $u^2 \le u_1^2$, when u_1^2 is the solution of the equation:

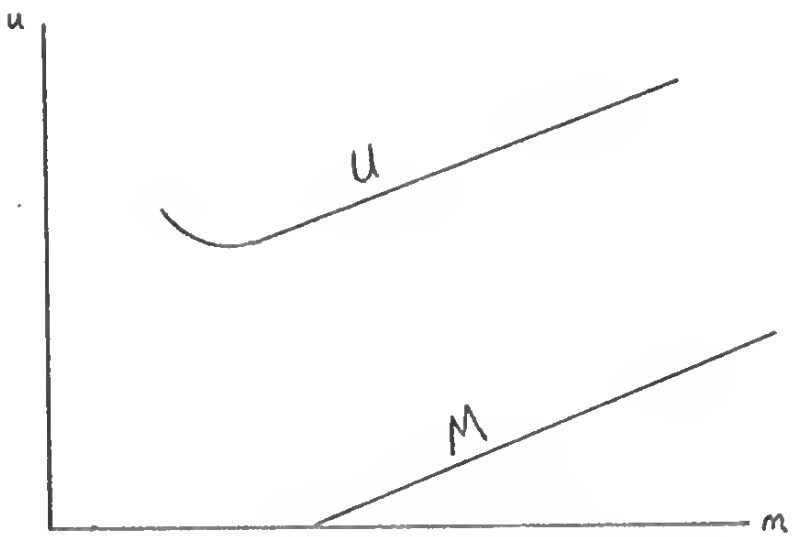
$$M(\tfrac{1}{2}n,\ \tfrac{1}{2},\ \tfrac{1}{2}\delta^2\ u_1^2) = \frac{\beta}{1-\alpha}\ e^{\frac{1}{2}n\delta^2} , \tag{6.9}$$

and H_1 is accepted as soon as $u^2 \geq u_2^2$, where u_2^2 is the solution of the equation:

$$M(\tfrac{1}{2}n,\ \tfrac{1}{2},\ \tfrac{1}{2}\delta^2 u_2^2) = \frac{1-\beta}{\alpha}\, e^{\frac{1}{2}n\delta^2} . \qquad (6.10)$$

Boundary values for u_1^2, and u_2^2, solutions to equations (6.9) and (6.10), have been given in (9) for a wide range of values of α, β, δ, and n.

In practice values of u^2 are plotted on charts similar to the figure 6.



NUMBER OF PAIRS

Fig. 6. SCHEMATIC REPRESENTATION OF AN OPEN DESIGN FOR NORMALLY DISTRIBUTED MEASUREMENTS WITH UNKNOWN VARIABILITY (SEQUENTIAL T-TEST).

Armitage (3) gave an example of open sequential t-tests to compare the relative efficacies of different analgesics for patients with rheumatoid arthritis.

One subject was treated for a week with treatment A and then, was treated for a week with treatment B. Three observations were recorded each week. The performance was measured by a score from a grip test. An average score was figured.

The next subject received the treatments in the order BA. The difference between the average grip of successive patients was used in calculating u^2. A critical difference δ was chosen to be .85, because this was believed to be the difference between aspirin and a placebo.

Three trials were reported. In the first trial prednisone was shown to be better than aspirin after 7 recordings. In the second no significant difference was detected in phenylbutazone and aspirin after 13 pairs of patients. The third report showed aspirin to be better than N-acetyl-para-aminophenol after 29 pairs.

ACKNOWLEDGEMENTS

I wish to thank Dr. S. Zacks, for his suggestion of this topic and for his advice and assistance during the preparation of this report.

I also wish to thank Dr. L. S. Laws for his encouragement and help in preparation for graduate school.

REFERENCES

1. Armitage, P. (1958). "Numerical Studies in the Sequential Estimation of a Binomial Parameter." Biometrika 44:1-15.

2. Armitage, P. (1957). "Restricted Sequential Procedures," Biometrika 44:9-26.

3. Armitage, P. (1960). Sequential Medical Trials. Blackwell, Oxford.

4. Armitage, P. (1963). "Sequential Medical Trials: Some Comments on F. J. Anscombe's Paper." J. American Statistical Association 58:384-387.

5. Bartlett, M. S. (1946). "The Large Sample Theory of Sequential Tests." Cambridge Philosophical Society Proceedings. 42:239-244.

6. Bross, I. (1952). "Sequential Medical Plans." Biometrics 8:188-205.

7. Choi, S. C. (1968). "Truncated Sequential Designs for Clinical Trials Based on Markov Chains." Biometrics 24:159-168.

8. Marshall, J. and Shaw, D. A. (1960). "Anticoagulant Therapy in Acute Cerebrovascular Accidents." Lancet i 995-998.

9. National Bureau of Standards (1951). Tables to Facilitate Sequential t-Tests, Applied Mathematics Series Number 7, prepared by K. J. Arnold.

10. Robertson, J. D. and Armitage, P. (1959). "Report of a Clinical Trial to Compare Two Hypotensive Agents." Anaesthesia 14:53-64.

11. Ruston, S. (1950). "On a Sequential t-Test." Biometrika 37:326-333.

12. Schneiderman, M. A. and Armitage, P. (1962). "A Family of Closed Sequential Procedures." Biometrika. 49:41-56.

13. Schneiderman, M. A. and Armitage, P. (1962). "Closed Sequential t-Tests." Biometrika 49:359-366.

14. Smith, J. Morrison (1958). "Hydrocortisone Hemisuccinate by Inhalation in Children with Asthma." Lancet ii 1248-1250.

15. Stockman, C. M. and Armitage, P. (1946). "Some Properties of Closed Sequential Schemes." J. R. Statist. Soc. Suppl. 8:104-112.

16. Wald, A. (1947). Sequential Analysis. Wiley, New York.

17. Wetherill, G. B. (1966). Sequential Methods in Statistics. Wiley, New York.

SEQUENTIAL PROCEDURES IN MEDICINE

by

JERRY WAYNE HIGHFILL

B. A., Southwestern College, 1967

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics
and Computer Science

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1968

Ethical considerations require a physician to protect as many patients as possible during a trial which compares different treatments. As differences in treatments become apparent, a medical person is required to use what he considers the best treatment. On the average sequential testing requires fewer observations than are needed in fixed sample procedures and respond rapidly when large differences are observed. Although some difficulties may be encountered in a design of this type, it is deemed particularly appropriate that sequential procedures be used in many investigations of medicine.

This report reviews some of the problems in designing sequential medical trials and presents some examples of experiments performed in actual clinical trials. The sequential probability ratio test is discussed. A lemma is included which proves that the sequential prohability ratio test terminates with probability one, and formulas for the average sample number and the operating characteristic function are developed. Procedures for working with data that is either qualitative or quantitative is derived.

Since an unexpected long series of observations is possible and may be very undesirable in a clinical trial, truncated and restricted designs are introduced for the binomial case. An exact method of counting sample paths by matrix multiplication and by using a fundamental equation of Markov chains is presented for the truncated designs; and an approximate method, the diffusion process, is discussed for restricted designs.